



Kwon, M., Kwon, H-H., & Han, D. (2020). A Hybrid Approach Combining Conceptual Hydrological Models, Support Vector Machines and Remote Sensing Data for Streamflow Simulation. *Remote Sensing*, 12(11), [1801]. <https://doi.org/10.3390/rs12111801>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.3390/rs12111801](https://doi.org/10.3390/rs12111801)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via MDPI at <https://doi.org/10.3390/rs12111801> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Article

A Hybrid Approach Combining Conceptual Hydrological Models, Support Vector Machines and Remote Sensing Data for Rainfall-Runoff Modeling

Moonhyuk Kwon ¹, Hyun-Han Kwon ^{2,*}  and Dawei Han ³ 

¹ K-water Research Institute, Korea Water Resources Corporation, Daejeon 34350, Korea; watermun@kwater.or.kr

² Department of Civil and Environmental Engineering, Sejong University, Seoul 05006, Korea

³ Civil and Environmental Engineering, University of Bristol, Bristol BS8 1TR, UK; d.han@bristol.ac.uk

* Correspondence: hkwon@sejong.ac.kr

Received: 18 April 2020; Accepted: 1 June 2020; Published: 2 June 2020



Abstract: Understanding catchment response to rainfall events is important for accurate runoff estimation in many water-related applications, including water resources management. This study introduced a hybrid model, the Tank-least squared support vector machine (LSSVM), that incorporated intermediate state variables from a conceptual tank model within the least squared support vector machine (LSSVM) framework in order to describe aspects of the rainfall-runoff (RR) process. The efficacy of the Tank-LSSVM model was demonstrated with hydro-meteorological data measured in the Yongdam Catchment between 2007 and 2016, South Korea. We first explored the role of satellite soil moisture (SM) data (i.e., European Space Agency (ESA) CCI) in the rainfall-runoff modeling. The results indicated that the SM states inferred from the ESA CCISWI provided an effective means of describing the temporal dynamics of SM. Further, the Tank-LSSVM model's ability to simulate daily runoff was assessed by using goodness of fit measures (i.e., root mean square error, Nash Sutcliffe coefficient (NSE), and coefficient of determination). The Tank-LSSVM models' NSE were all classified as "very good" based on their performance during the training and testing periods. Compared to individual LSSVM and Tank models, improved daily runoff simulations were seen in the proposed Tank-LSSVM model. In particular, low flow simulations demonstrated the improvement of the Tank-LSSVM model compared to the conventional tank model.

Keywords: rainfall-runoff; support vector machine; hybrid model; satellite soil moisture

1. Introduction

A hydrologic model is a simplified representation of a complex system that facilitates the understanding of the hydrologic cycle in a simplified manner. The rainfall-runoff (RR) model plays a central role in describing different aspects of water resources management. These include streamflow record extension for designing hydraulic structures [1], operational applications to hydraulic structures [2–4], streamflow prediction for ungauged catchments [5], and assessments of land use and climate change impacts [6,7]. Over the last three decades, the hydrology field has focused on the development and evaluation of runoff prediction models. RR models have undergone considerable efforts and improvements to understand the impacts of rainfall events on catchment response, but whether these improvements meet scientific and practical demands is still in question [8]. Spatio-temporal variability is substantially dependent on spatial aspects. That is, the runoff response to rainfall is intricately linked to influencing factors such as climate conditions; land cover; soil properties; topography; and human activities, including irrigation and urbanization [6,9,10].

RR models can be categorized into three main groups: data-driven, conceptual, and physically-based models [10]. Data-driven models, also known as empirical or black-box models, build RR relationships directly using relatively large datasets of historical records, statistical models, and machine learning schemes. In other words, data-driven models statistically represent the RR relationships that govern catchment response to rainfall and can be used to predict runoff based on a set of predictors over time [11,12]. In contrast to data-driven models, conceptual models consist of a series of interconnected systems that represent hydrological processes at spatial and temporal scales [13,14] and physically-based models are formulated in terms of mathematical relationships and hydrologic process interactions that describe how rainfall transforms into runoff [12,15].

RR modeling began with conceptual models from the 1960s to the 2000s, followed by recent advances using data-driven approaches, mainly through the use of machine learning techniques. Conceptual models, such as the Sacramento model [16], IHACRES [6], PDM model [17], and HBV model [18], typically require fewer parameters compared to physically-based models (e.g., MIKE-SHE [19], SWAT [20], and TOPMODEL [15]). This leads to less computational complexity and provides near real-time predictions. Conceptual models are relatively straightforward to construct and implement and can provide comparable estimates of runoff at lower computational costs than physically-based models [8,21]. According to [8], the degree of complexity and sophistication of the hydrological model does not ensure an improvement in performance; performance is more likely to rely on hydrological variables (e.g., rainfall, runoff, and temperature) or many other state variables, including the initial catchment wetness. The tank model proposed by [22] is popular with the hydrological community primarily due to its simple structure and reasonable accuracy, without considering soil moisture (SM) information. In this respect, this well-tested model is effective and useful for both practical and theoretical aspects of RR modeling [13,14,23,24]. The tank model is used for daily streamflow simulations over the entire country of South Korea to support long-term water resource management and planning every 5 to 10 years. For these reasons, we selected the tank model for our study.

For the past few decades, machine learning techniques have been widely used for hydrological modeling (e.g., artificial neural networks (ANN), SVM, fuzzy logic models, and genetic algorithms). Among these methods, SVM has been used in solving hydrologic classification problems [25] and predicting runoff ([26–28], among others). Data-driven modeling combines machine learning schemes with existing conceptual and physically-based models to provide a platform for RR modeling. Rather than using a nonlinear regression model, [3] combined a singular spectrum analysis with the SVM to provide improved accuracy to the non-linear prediction (NLP) method. [12] proposed a hybrid model by combining the Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS) and two different types of machine learning techniques (i.e., ANN and SVM). Other approaches that combine existing RR models with machine learning techniques can be found in the literature (e.g., [29,30]). In this study, we used the least squares support vector machine (LSSVM). The LSSVM has a computational advantage over conventional support vector machines by converting a quadratic form into a set of linear equations [9,31]. The LSSVM is also known as a nonlinear model, particularly when used for solving classification and regression problems such as time-series predictions.

Apart from the model structure, SM is a key variable in the interaction between surface rainfall and infiltration. SM can be attributed to hydrological elements such as evapotranspiration, infiltration, and percolation loss [32,33]. Runoff modeling is highly dependent on how accurately the RR model captures SM's spatio-temporal variations [34,35]. Despite the importance of SM's spatio-temporal variations, its use in hydrological models is limited by the lack of in situ observation networks. Satellite-based SM products have become available over the last decades with increased accuracy and frequency [34]. Their hydrological applications have expanded and are used as alternative data [36], rainfall estimation [37,38], and drought monitoring [39]. In addition to the aforementioned applications, satellite SM retrievals have been integrated into conceptual [32,40] and physically-based RR models [12,34,41] using data assimilation techniques. However, the effect of SM on the RR model

varies greatly depending on the study, providing either no or limited improvement on the accuracy of the simulation. For example, [41] found that SM made a limited contribution to runoff prediction, while [40,42] achieved a considerably improved overall response by combining satellite SM data with a conventional RR modeling framework. Conversely, no measurable improvement was found by [43].

Given this background, this study attempts to address the following questions:

1. How much do the intermediate variables obtained from the RR model correlate with in situ SM or remotely sensed SM products?
2. Can the intermediate variables be effective in RR modeling within a machine learning-based regression framework, particularly for low flow simulations in a hybrid model?

This study investigates a hybrid RR model that uses intermediate state variables obtained from a conceptual model (i.e., the tank model) within an LSSVM-based regression framework. The fundamental hypothesis behind the proposed hybrid approach is that the combined use of different models may capture complex features of the RR process, thus providing a better representation than individual models. First, we built the tank model and explored its performance in terms of simulating daily runoff discharge in the Yongdam Catchment, South Korea. We assumed that the water depths at each tank were the intermediate state variables to be considered as a proxy measure of SM. We then explored the water depths at each tank and their role in the LSSVM-based RR model. In addition, we examined the potential use of remotely sensed SM products and combined them with the state variables.

The hydrologic and satellite data used in this study are illustrated in the following section. The theoretical backgrounds for the RR model and the machine learning model are presented in Section 3. A detailed discussion of the proposed modeling scheme is provided in Section 4, and this paper ends with a summary of the findings and future work in Section 5.

2. Study Area and Data

2.1. Study Area and In Situ Observations

The Yongdam Catchment has a drainage area of 930 km². The primary land use consists of forest, paddy, and mountainous areas, covering 71.1%, 9.4%, and 9.9% of the catchment area, respectively. An annual rainfall of 1299 mm and runoff depth of 680 mm was measured in the area between 2007 and 2016. The Yongdam Multipurpose Dam, operated by Korea Water Resources Cooperation (K-water), was built in 2000 and used for water supply (1,050,000 m³/d), hydroelectric power (24,400 kW), and flood control [44]. Hydrologic data such as runoff, rainfall, and SM records are measured by K-water (<http://english.kwater.or.kr>). There are six hydrologic stations where precipitation has been measured since 2001. Here, the Yongdam dam inflow records are regarded as the unregulated flow (or natural flow). SM has been recorded since 2014 using a time-domain reflectometer (TDR; [45]). The study area and the location of stations used in this research are presented in Figure 1. Here, the areal rainfall is defined by Thiessen Polygons, and climate data (i.e., air temperature, relative humidity, wind speed, and hours of sunshine) used to estimate reference evapotranspiration were obtained from the Jangsu weather station operated by the Korea Meteorological Administration (KMA; <https://web.kma.go.kr/eng/>).

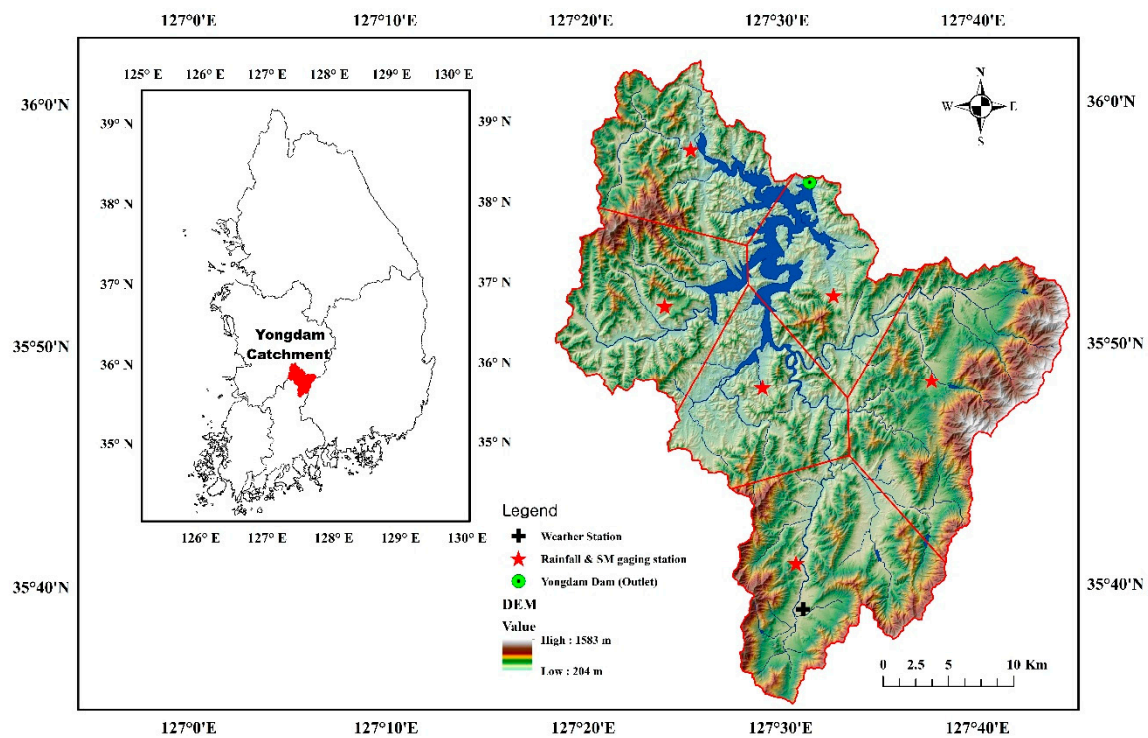


Figure 1. Map showing the study area along with the location of hydrologic and weather stations used in this research. The polygon enclosed within solid lines represents a Thiessen polygon for estimating the areal rainfall over the Yongdam Dam watershed. Note that the discharge gauging station is located at the Yongdam dam site.

2.2. Satellite SM Measurements

The European Space Agency (ESA) released satellite-based SM products on a global scale by combining active and passive microwave sensors [33]. The SM products from the individual sensors were blended by spatiotemporal resampling and rescaling (i.e., the cumulative distribution function matching) techniques [46]. The ESA provides SM datasets from 1978 to 2016 to the public domain with different units (i.e., volumetric units (m^3/m^3) for the passive and combined product, and percent of saturation units (%) for the active dataset). The active measurements were obtained by merging Active Microwave Instruments Windscat (AMI-WS) and ASCAT. Passive products were acquired by combining several remote sensing data sources, including SMMR, SSM/I, TRMM Microwave Imager (TMI), AMSR-E, WindSat, AMSR2 (successor to AMSR-E) and SMOS data. Additional details on the methodology and comparison to the ESA CCI SM can be found in the literature [33,46,47]. The ESA CCI SM products (v04.2) are used in this study and were obtained from their website (<https://www.esa-soilmoisture-cci.org/>; accessed and downloaded on 17th September 2018). The active-passive combined SM products with a 0.25° spatial resolution were chosen for this study due to their higher temporal resolution and better accuracy compared to both the active and passive retrievals [48]. It should be noted that both the satellite soil moisture retrievals and in situ soil moisture were averaged over the entire watershed for a representation. To be specific, the satellite pixel values whose centroids are located in the study site are extracted then averaged. Similarly, daily in situ soil moisture data are first collected then averaged to reproduce a catchment mean soil moisture time series.

3. Methodology

3.1. The Tank Model

The tank model is classified as a deterministic, conceptual, and continuous model. Similar to other conceptual models consisting of a series of interconnected subsystems (i.e., storages), the tank model is

composed of three vertically interconnected tanks (e.g., 3-tank), as illustrated in Figure 2. The structure of the tank model, such as the number of tanks and their associated side outlets, can be formulated in terms of physical catchment attributes and climatological conditions. For instance, a model with two tanks (e.g., 2-tank) was used to evaluate the RR relationship for a paddy field [23], while [14] used a 4-tank model to represent the deep percolation process in a forested region.

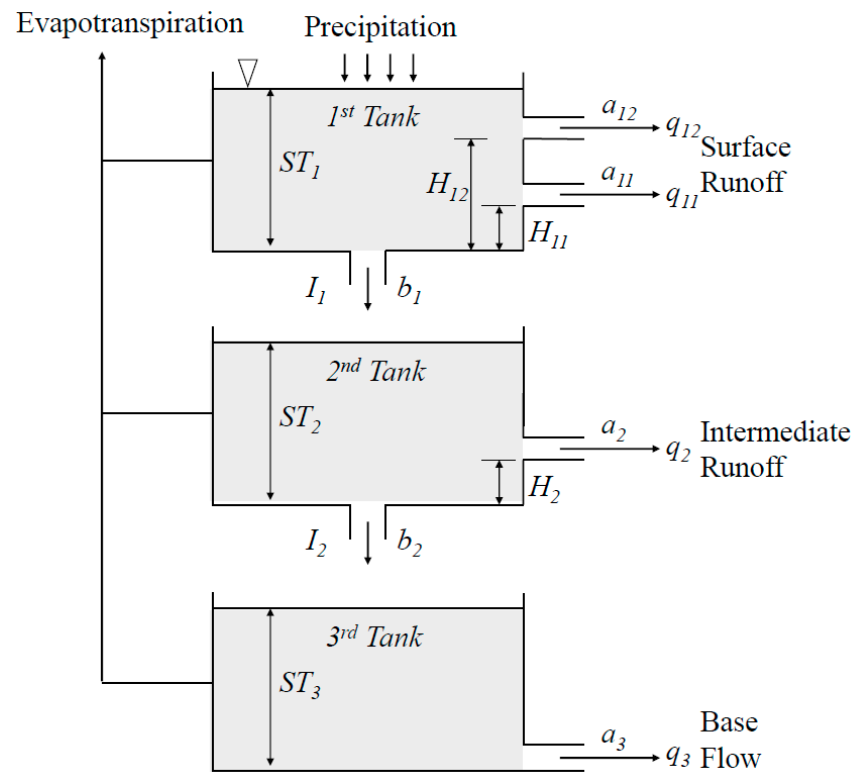


Figure 2. Schematic diagram showing the structure of the 3-tank model.

The 1st tank's level is determined by rainfall, while the other two tanks' depths are governed by infiltration and percolation. The tanks are depleted through evapotranspiration and runoff, and the amount of water left in each tank represents catchment storage. A continuous daily rainfall is used as input for the 1st tank; evapotranspiration is assumed to occur in all three tanks. If there is insufficient water in the 1st tank, the storage level in the 2nd tank is reduced to fulfill the demand for evapotranspiration in the 1st tank. Similarly, a lack of water in both of the top two tanks affects storage in the 3rd tank. The side outlets in each tank simulate runoff in different layers, including surface runoff, intermediate runoff, and baseflow. The total runoff is calculated by a summation of the runoff accumulated from the side outlets as follows [13]:

$$Q = q_{11} + q_{12} + q_2 + q_3 = \sum (ST_i - H_{ij})a_{ij}, \quad (1)$$

where Q is total runoff and q_{ij} refers to the runoff of j^{th} side outlet at the i^{th} tank (mm) with the associated runoff coefficient a_{ij} . ST_i represents the storage of the i^{th} tank (mm). H_{ij} is the height of side outlet for the j^{th} side outlet in the i^{th} tank (mm). More details on the model equation can be found in Appendix A.

To accurately estimate runoff with the tank model, reliable daily rainfall and evapotranspiration datasets over a given catchment area are required for the input data. In our study, we used daily rainfall sequences from six stations and the areal mean rainfall over the catchment area calculated by the Thiessen polygon method. The standard FAO-56 Penman–Monteith method was used to estimate the reference evapotranspiration [49]. The potential evapotranspiration estimates were adjusted via

the model's calibration process and considering the water balance over the catchment between the input (rainfall) and output (runoff).

3.2. The Least Square Support Vector Machine (LSSVM) Model

The SVM developed by [50] has been widely used for classification and regression tasks in many different fields, including RR predictions [31,51]. The LSSVM is a modified version of the SVM. More specifically, the LSSVM differs from the conventional SVM in that the LSSVM approach uses a set of linear equations for solving optimization problems instead of the quadratic form used in the conventional SVM [9,52]. The LSSVM is presented schematically in Figure 3. The LSSVM adopts a least squares linear system as the loss function with two-layer networks. Below is a brief description of the LSSVM; a more detailed explanation can be found in the literature [53].

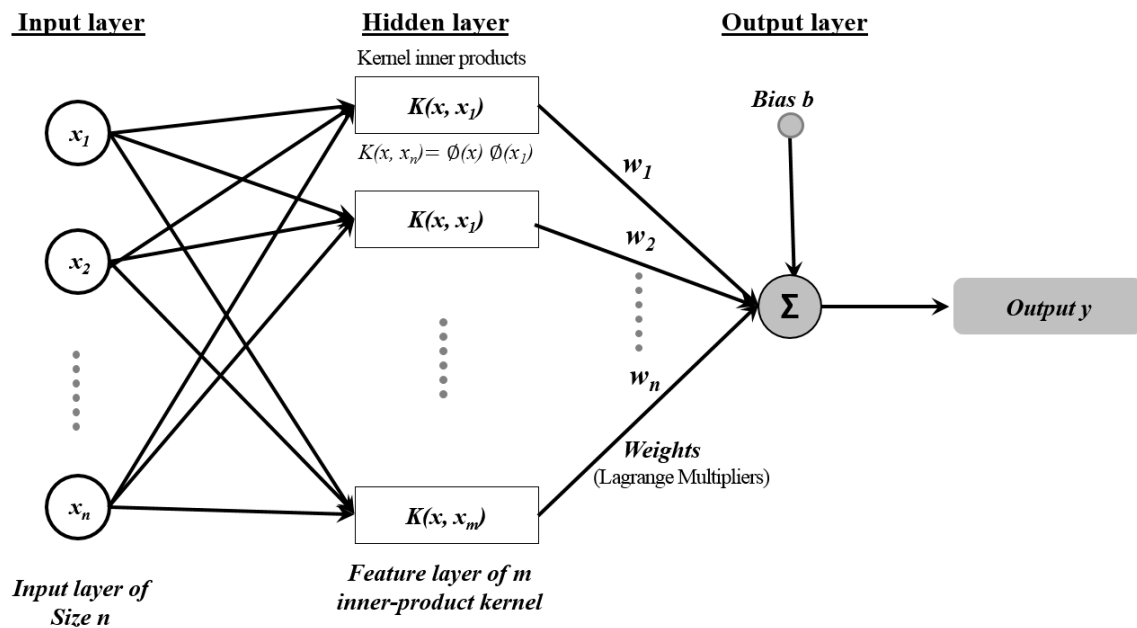


Figure 3. Schematic architecture of the SVM based on a radial basis functions (RBF) network.

Let $\{x_k, y_k\}_{k=1}^N$ be an N length set of datasets with the input ($x \in R^N$) and output ($y \in R$), where R^N denotes the N -dimensional input/output vector space. The LSSVM model can be formulated as follows:

$$y(x) = w^T \varphi(x) + b, \quad (2)$$

where $\varphi(\cdot)$ is the feature map embedding the input data into a higher-dimensional feature space, w is a weight vector, and b is a bias term.

For the LSSVM regression, the optimization problem is given as follows:

$$\begin{aligned} \text{Minimize : } & \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2, \\ \text{Subject to : } & y(x) = w^T \varphi(x_k) + b + e_k, k = 1, \dots, N, \end{aligned} \quad (3)$$

where γ is the regulation parameter and e_k^2 is the error variable. Equation (3) can be rewritten by introducing Lagrange multipliers as follows:

$$y(x) = \sum_{k=1}^N a_k K(x, x_k) + b, \quad (4)$$

where a_k is the Lagrangian multiplier and $K(x_k, x_l)$ is the kernel function.

$$K(x_k, x_l) = \exp\left\{-\frac{(x_k - x_l)^T(x_k - x_l)}{2\sigma^2}\right\}, k, l = 1, \dots, N \quad (5)$$

σ is the width of the radial basis function.

The LSSVM's performance is highly dependent on the kernel function. Among the many kernel functions, such as linear, polynomial, sigmoidal, and radial basis functions (RBFs), an RBF was used for its practical aspects. The RBF is more flexible than the others and has fewer parameters to estimate [30]. The LSSVM with a Coupled Simulated Annealing (CSA) optimization algorithm [54] determines an initial set of suitable parameters based on five multiple starters. These parameters are used in the second optimization procedure to further fine-tune the parameters. A simplex optimization approach [55] is then employed here for tuning the parameters through a cross-validation procedure by partitioning samples into training data and test data. To compare the tank model under the same conditions, hydrologic data used for the calibration period (2007–2013) were also used for the LSSVM model's training process. The remaining verification period data (2014–2016) were used for the testing phase.

3.3. The Tank-LSSVM Hybrid Model

Our hybrid RR model, hereinafter called the Tank-LSSVM model, was constructed using the output (i.e., ST_1, ST_2, ST_3) obtained from the tank model in an LSSVM-based regression framework. In this study, the tank model's storage variables ST_1, ST_2 , and ST_3 were considered intermediate state variables representing the SM temporal variation over the catchment area. A schematic representation of the proposed hybrid model, together with specific implementation procedures, is presented in Figure 4. Input variable (or predictor) selection is important in a machine learning-based approach [26]. In this study, predictors considered for the Tank-LSSVM model were rainfall, tank storage, and ESA CCI SM data. Satellite SM products are of particular interest for ungauged watersheds. We conducted an experimental study to explore the use of satellite SM products within the proposed model structure. We used an iterative approach to obtain the optimal combination of independent variables in a time-lagged stepwise regression manner by minimizing the difference between the simulated and observed flow. The partial autocorrelation function (PACF) was utilized to explore the relative importance of the time-lagged variables in the hybrid prediction model.

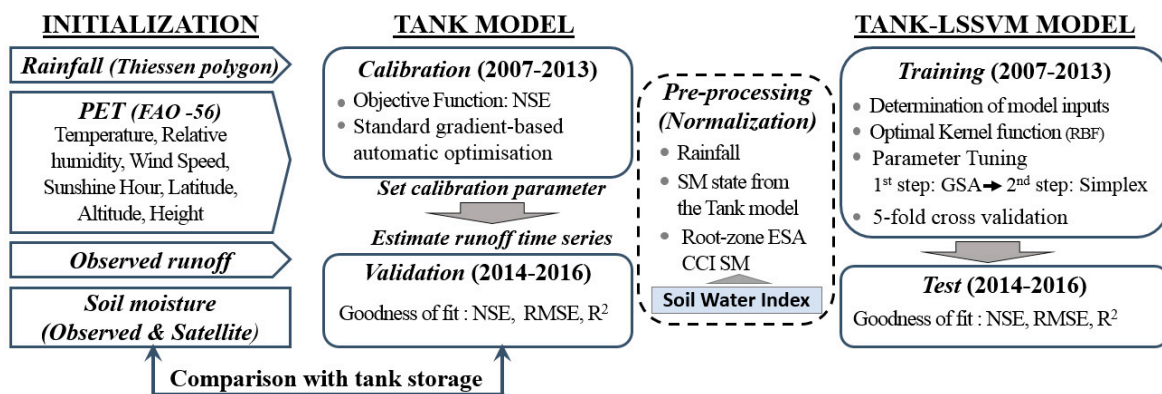


Figure 4. Schematic representation of the Tank-least squared support vector machine (LSSVM) modeling process.

Prior to the construction of the LSSVM-based RR model, the exploratory (independent) variables considered in this study were all normalized. This is a common approach in data-driven models. It reduces the problems caused by relatively high values [56]. In this context, normalization is done by

subtracting the mean (\bar{x}) from each variable (x) prior to dividing by the standard deviation (S). This is called a Z-score (Z).

3.4. Root-Zone ESA CCI SM Products

One fundamental issue in using remote sensing SM products is that SM is measured a few centimeters below the surface. This makes the data difficult to use in hydrological applications [57]. Root-zone SM contents are more representative when simulating a catchment response to rainfall [32,35]. Given the mismatch between the satellite SM measurements and the tank model's configuration for soil layers, the exponential filter method, also known as the soil water index (SWI), was used to derive the root-zone SM from the original ESA CCI SM data. This approach is a common preprocessing step for satellite-derived SM data used as an input to hydrological models [34,35,58,59]. In this study, a recursive formulation of the exponential filtering method introduced by [60] was adopted and expressed as:

$$SWI_n = SWI_{n-1} + K_n [SSM_{(t_n)} - SWI_{n-1}], \quad (6)$$

where SWI_n and $SSM_{(t_n)}$ are the soil water index and the ESA CCI SM at t_n , respectively. The gain parameter K ranging from 0 to 1 can be obtained from the following relationship:

$$K_n = \frac{K_{n-1}}{K_{n-1} + e^{-\frac{(t_n - t_{n-1})}{T}}}. \quad (7)$$

The filtering was initialized by applying $SWI_0 = SSM_{(t_0)}$ and $K_0 = 1$ in Equation (6). The optimal characteristic time length T was obtained by maximizing the correlation coefficient (r) between the SWI and simulated SM (i.e., tank storage) from the tank model. The parameter T substantially adjusted the SM temporal variation. More specifically, this approach smoothed the original ESA CCI SM surface series and could be considered an SM proxy for a deeper layer. For further details, the reader is referred to [60,61].

3.5. Performance Scores

The proposed RR model's efficiency was evaluated using three goodness-of-fit (GOF) measures—the Nash–Sutcliffe coefficient, the coefficient of determination, and the root mean square error. These are commonly used in hydrologic and hydroclimatic models [62]. A more detailed description of the GOF measures is summarized in Table 1. In this study, we adopted RMSE_Q70 to quantify the model performance, particularly for low flows within the range of 70–100% time exceedance, following [7]. We also considered the descriptive performance criteria, using the NSE as proposed by [63] to determine the degree of accuracy in simulating the daily runoff. The performance criteria were very good: $NSE \geq 0.7$; good: $0.5 \leq NSE < 0.7$; satisfactory: $0.3 \leq NSE < 0.5$; unsatisfactory: $NSE < 0.3$.

Table 1. Performance metrics employed in this study. O and \bar{O} indicate the observed runoff and observed runoff mean, respectively. E is the simulated runoff and \bar{E} is the simulated runoff mean. n is the number of observations.

Performance Metrics	Equations	Range	Optimal Value
Nash–Sutcliffe efficiency (NSE)	$NSE = 1 - \frac{\sum_{i=1}^n (O_i - E_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$	$-\infty \sim 1$	1
Coefficient of determination (R^2)	$R^2 = 1 - \frac{[\sum_{i=1}^n (O_i - \bar{O}) \cdot (E_i - \bar{E})]^2}{\sum_{i=1}^n (O_i - \bar{O})^2 \cdot \sum_{i=1}^n (E_i - \bar{E})^2}$	$0 \sim 1$	1
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - E_i)^2}{n}}$	$0 \sim \infty$	0

4. Results and Discussion

4.1. Rainfall-Runoff Using the Tank Model

The tank model is a conceptual RR model that requires a small amount of data and relatively few model parameters. Through a model calibration process, ten parameters were adjusted to approximate the observed streamflow, and the validation process was carried using the calibrated parameters. The purpose of the validation process was to ensure that the model could be used with new data and successfully reproduce daily streamflow observations over different periods of time. In this context, both the calibration and validation were performed under different climate regimes to evaluate the model performance better. In this study, daily rainfall and streamflow data from 2007 to 2013 were used for calibration because this date range covers average, wet, and dry climate conditions. Data from 2014 to 2016 were used for the model validation (Figure 5).

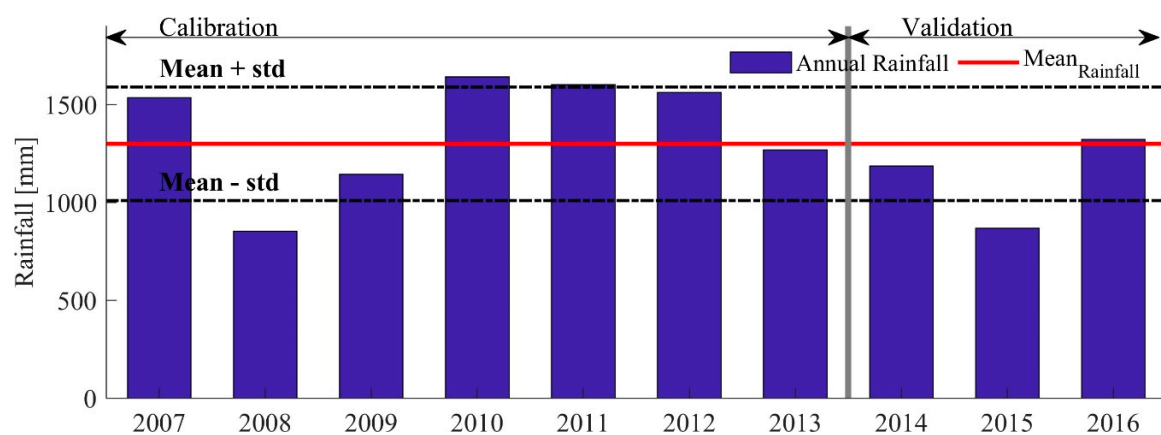


Figure 5. Temporal distribution of annual rainfall over the Yongdam catchment during the calibration and validation phases of this study. The red solid line and black dotted lines represent the mean and 1 standard deviation, respectively.

The optimal parameter set for the calibration period was derived from a standard gradient-based automatic optimization scheme [64], the MATLAB function ‘fmincon’, which is a built-in function in MATLAB. More details on the optimization procedure adopted in this study can be found at <https://www.mathworks.com/help/optim/ug/fmincon.html>. The model parameters and their values, determined via the optimization scheme, are summarized in Table 2. Figure 6 shows the daily simulated and observed runoff (Q) together with rainfall recorded during the 2007–2016 investigation period. The results suggest that the tank model’s performance for both the calibration and validation periods can be regarded as “very good” in terms of the NSE. The results show that the runoff process is reproduced effectively by the tank model, with an $R^2 = 0.92$ (0.81) and $RMSE = 20.18 \text{ m}^3/\text{s}$ ($14.72 \text{ m}^3/\text{s}$) for the calibration and validation phases, respectively. However, the tank model does not fully capture the RR process’s complex behavior in the validation phase compared to the calibration period (i.e., 0.92 for the calibration period and 0.81 for the validation phase). A notable difference between the observed and simulated runoff is detected during low flow periods. This result suggests that the tank model alone may be limited in its ability to describe low flow dynamics adequately. A more rigorous approach to low flow simulation is to use a hybrid model that incorporates the intermediate variables obtained from the tank model into the LSSVM-based RR modeling framework.

Table 2. Parameter ranges and optimum values estimated through the calibration process. Here, the parameter ranges were set based on previous studies (Kim and Park, 1988; Song et al., 2017).

Parameter		a11	a12	a20	a30	b1	b2	b3	h11	h12	h20
Range	Min.	0.08	0.08	0.03	0.00	0.10	0.01	0.00	5.00	20.00	0.00
	Max.	0.50	1.00	1.00	0.03	0.50	0.35	0.11	60.00	150.00	100.00
Obtained value		0.13	0.33	0.71	0.02	0.14	0.07	0.01	10.72	62.94	35.14

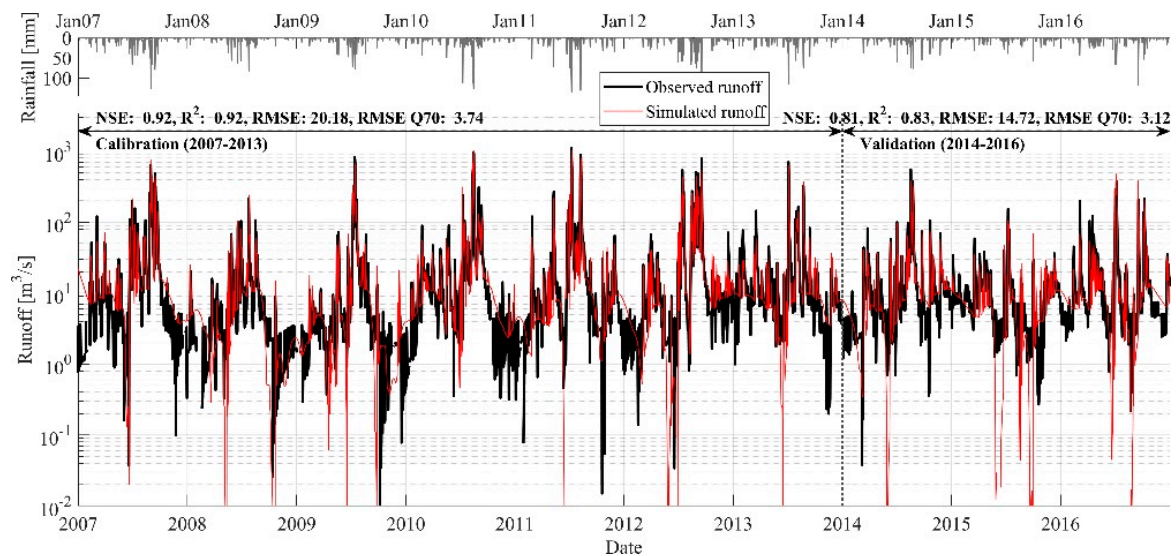


Figure 6. Time series plot showing the observed runoff and the runoff simulated by the tank model. The upper panel depicts the average catchment rainfall over the course of the investigation period.

First, we explored the correlation between tank storage and in situ SM to determine if the tank storage was an effective proxy for SM and if it represented the SM temporal variability over the catchment area. Six cases were studied to understand how the individual tanks (i.e., Cases 1-3) and their combinations (i.e., Cases 4-6) correlated with the in situ SM (Figure 7). The in situ SM observations were limited to relatively small periods between 2014 and 2016. We analyzed the lagged windowed cross-correlation to quantify the temporal coherence between the tank storage and in situ SM (Figure 8) and found a statistically significant correlation at the 95% confidence level. A confidence interval is obtained by the Gaussian distribution $N(0, 1/N)$, whose standard deviation is $1/\sqrt{N}$. For a 95% confidence level, the confidence interval can be defined as $CI = 0 \pm 1.96/\sqrt{N}$. The strongest temporal coherence for the zero-lag correlation between the individual tank storage and in situ SM was observed in Case 2 (the 2nd tank), where the correlation coefficient (r) was 0.77. The lowest r value (0.43) was found for Case 3, suggesting that the intermediate variables from each tank describe the SM temporal dynamics. This is because the upper layer of the soil is subjected to more rapid drying and rewetting, and soil moisture variations in this layer are more prominent compared to that of the lower layer. The tank combinations (i.e., Cases 4-6) demonstrated slightly higher or lower r values compared to Case 2. Case 2 had the highest time lag r value (0.78) with lag-1, while the strongest correlation for Case 3 was observed with lag-11. This may be due to the SM's slow response behavior in Case 3, since it represents the deepest soil layer in the tank model. Overall, the results obtained from the cross-correlation analysis confirm that the tank storage temporal dynamics are closely related to SM variations. This suggests that the tank model's intermediate variables can be used as a proxy for the SM.

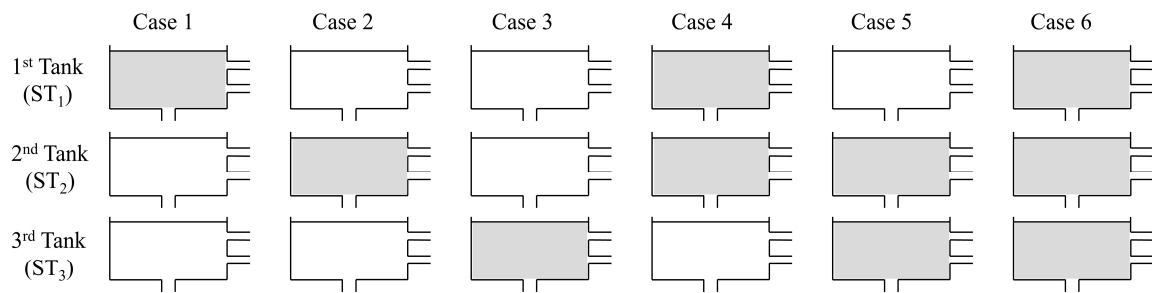


Figure 7. Schematic representation of six different tank combinations. Grey-colored tanks refer to the tanks used in comparison with in situ soil moisture (SM) observations.

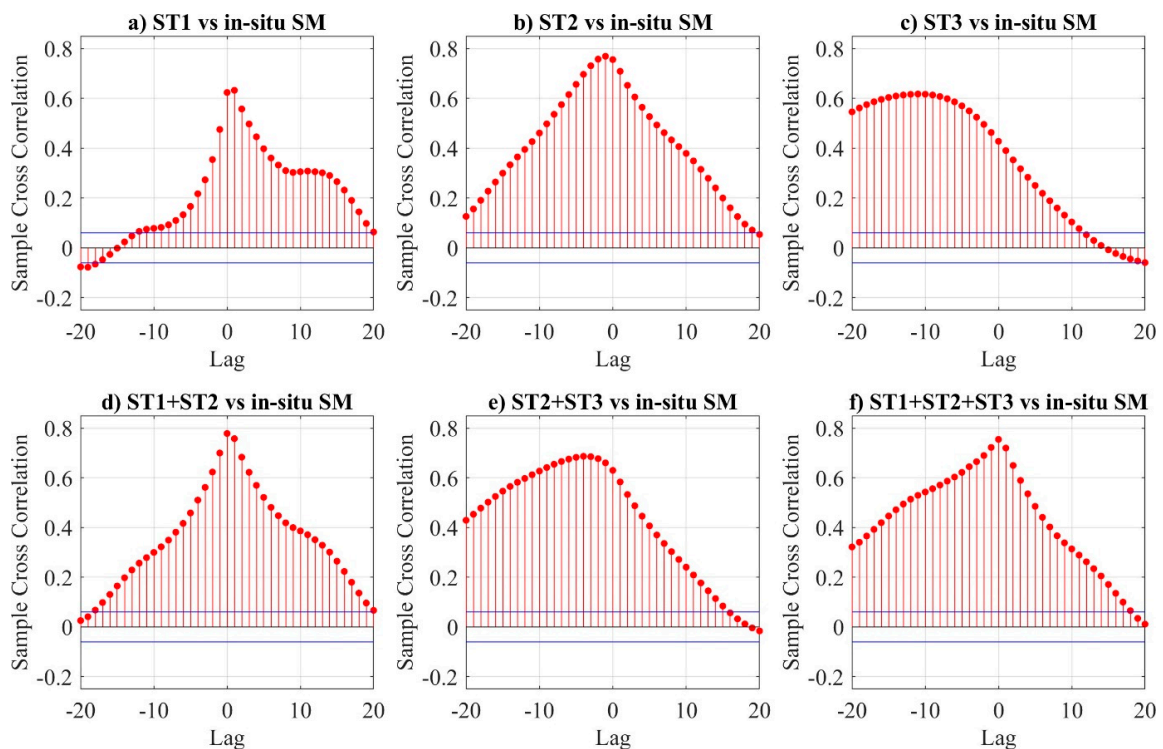


Figure 8. Lagged cross-correlation between the tank storage and in situ observations from 2014 to 2016. The solid blue line represents autocorrelation along with the 95% confidence interval for a white noise process. The values outside of the confidence interval are statistically significant.

4.2. LSSVM and Tank-LSSVM Models

4.2.1. Determination of Model Inputs

A machine learning-based hydrological model's performance is dependent on the lagged input vectors chosen for the model training [26,56]. Apart from rainfall and satellite SM, the tank storage derived from the tank model in this study is considered a proxy for the SM content. Additionally, the time-lagged relationships between the hydrologic variables enable the consideration of a sequential hydrological process in the proposed framework. Even though the input variables and time lags are important, their selection procedures are rather ad hoc when using the machine learning-based regression approach [26]. There is no universal or generalized model that selects the optimal time-lagged input vector directly; this is why an empirical approach is favorable. In this study, we used the partial autocorrelation function (PACF) to understand the smallest lag time for a parsimonious model. The PACF function analysis plots for each input variable are presented in Figure 9. Based on these results, the lag extent for each variable was determined and the input vectors were considered

predictors for the proposed hybrid model. For example, a statistically significant PACF in ST1 was observed by lag-2, and the PACF falls outside of the 95% confidence interval at lag-3. Here, the ST1 lagged vectors ranging from lag-0 to lag-2 were considered primary. Other variables were similar.

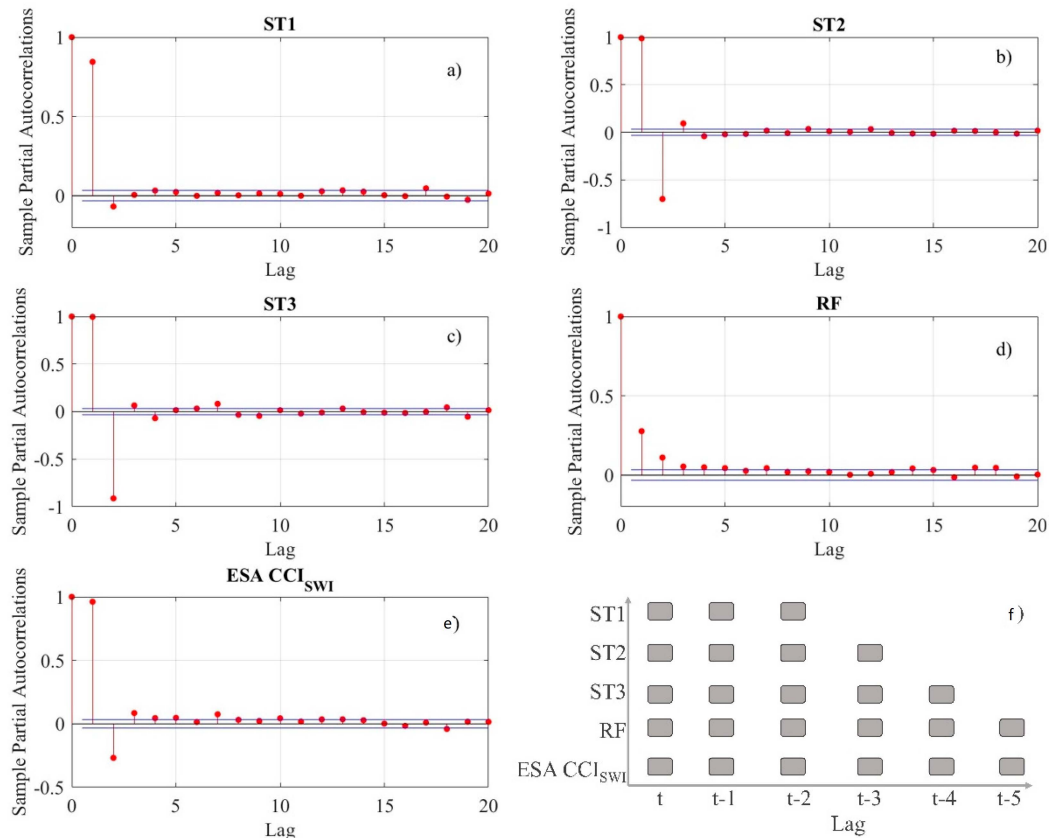


Figure 9. The partial autocorrelation function of the input variables (a–f) and their time lags considered in this study (f). ST_n, RF, and ESA CCI_{SWI} represent the n-th tank storage in the tank model, rainfall, and root-zone European Space Agency (ESA) CCI SM data, respectively. The solid blue line represents partial autocorrelation along with the 95% confidence interval for a white noise process. Values outside of the confidence interval are statistically significant.

4.2.2. LSSVM Model

The LSSVM-based RR model was constructed using several lagged independent variable values, rainfall (P_{t-n}) and ESA CCI_{SWI} (θ_{t-n}), as inputs without a set of intermediate variables from the tank model. The input vectors were partitioned into two subsets, a training phase and a testing phase, with the same period as the tank model to facilitate comparisons with the tank model results under the same conditions. Both the dependent and independent variables were normalized before applying the LSSVM model.

We built the first LSSVM-based runoff model using a single rainfall data predictor. The ESA CCI_{SWI} data were added as an additional input to explore the satellite-based SM contribution to the LSSVM-based RR model. Among many hydrological variables, rainfall and satellite-based SM inputs were selected specifically because the two variables are both closely related to the runoff process and also closely related. This allowed us to consider the interdependence in the hydrological cycle indirectly. Table 3 presents the results of the runoff simulation with different combinations of input variables. The relative impact of these combinations on the model performance was investigated in a stepwise manner by repeatedly adding more lagged values until the improvement stopped. To avoid overfitting, we mainly focused on performance improvement for unseen data during the testing phase.

Table 3. Model performance measures with different combinations of lagged input variables during the training and testing periods.

Model	Input Combinations	Training (2007–2013)				Testing (2014–2016)			
		NSE	R^2	RMSE	RMSE Q70	NSE	R^2	RMSE	RMSE Q70
SV1	$P_{(t)}$	0.60	0.60	44.72	10.40	0.40	0.49	26.02	9.86
SV2	$P_{(t)}, P_{(t-1)}$	0.84	0.84	28.21	6.73	0.45	0.61	24.81	5.42
SV3	$P_{(t)}, \dots, P_{(t-2)}$	0.88	0.88	24.88	5.11	0.57	0.65	22.04	4.49
SV4	$P_{(t)}, \dots, P_{(t-3)}$	0.89	0.89	23.01	4.44	0.62	0.70	20.77	3.99
SV5	$P_{(t)}, \dots, P_{(t-4)}$	0.91	0.91	21.33	4.00	0.68	0.75	18.91	3.86
SV6	$P_{(t)}, \dots, P_{(t-4)}, \theta_{(t)}$	0.91	0.91	21.24	3.18	0.69	0.77	18.80	4.79
SV7	$P_{(t)}, \dots, P_{(t-4)}, \theta_{(t)}, \theta_{(t-1)}$	0.91	0.91	20.74	3.29	0.69	0.77	18.62	5.02
SV8	$P_{(t)}, \dots, P_{(t-4)}, \theta_{(t)}, \dots, \theta_{(t-2)}$	0.92	0.92	19.53	3.12	0.72	0.79	17.90	4.93
SV9	$P_{(t)}, \dots, P_{(t-4)}, \theta_{(t)}, \dots, \theta_{(t-3)}$	0.92	0.92	19.33	2.98	0.73	0.81	17.51	4.75

An obvious increase in the NSE performance efficiency was seen (0.40 for the SV1 and 0.68 for the SV5) until lag-4, after which no further performance improvement was seen. Similar results were obtained for other performance measures, such as an increase in R^2 from 0.49 to 0.75 and a decrease in RMSE from 26.02 to 18.91 m³/s. Therefore, rainfall data from lag-0 to lag-4 were used in the subsequent analysis. As summarized in Table 3, the SV6–SV9 models introduced the lagged ESA CCI_{SWI} data as an input to the LSSVM modeling framework and showed better, more comparable performance than the SV1–SV5 models. The only exception to this was RMSE Q70 during the testing period, which indicated that the SM states inferred from the ESA CCI_{SWI} provided an effective means of describing the SM's temporal dynamics in the LSSVM-based RR model. A slight increase in the NSE values, ranging from 0.69 to 0.73, was identified for the SV6–SV9 models. Similar improvements in the performance measures R^2 , RMSE, and RMSE Q70 were also observed. Based on the performance criteria, the SV8 and SV9 models demonstrate “very good” performance (NSE > 0.7). No improvement in RMSE Q70 was seen when the ESA CCI_{SWI} data was included, suggesting that the direct use of remote sensing SM products played a limited role in the simulation of low flow. The simulated runoff was compared to the observed runoff over the course of the testing period and depicted as scatter plots with the corresponding R^2 (Figure 10), where enhanced results are seen for the SV6–SV9 models.

4.2.3. Tank-LSSVM Model

The Tank-LSSVM model bridges the intermediate state variables from the tank model and the machine learning framework (LSSVM). Several lagged values of tank storage (ST_{t-n}) derived from the tank model were utilized together with the rainfall and ESA CCI_{SWI} data in the LSSVM model. The RR simulation results from different input variable combinations are summarized in Table 4. The HY1–HY4 models were carried out without time lag consideration, whereas the HY5–HY9 models included time-lagged input vectors. Contrary to the LSSVM model results, the use of time-lagged input variables showed little or no improvement in the Tank-LSSVM model's performance when the intermediate SM state variables from the tank model were used (models HY5–HY9). This is partially due to the fact that the SM state variables derived from the well-tuned RR model themselves could represent the behavior of catchment-scale SM in this study without time-lagged values. The HY1 model only considered the 1st tank storage and rainfall as input variables, while the 2nd and 3rd tanks were included in the HY2 and HY3 models, respectively. The HY4 model combined the ESA CCI_{SWI} datasets with the HY3 model to assess the satellite SM contribution.

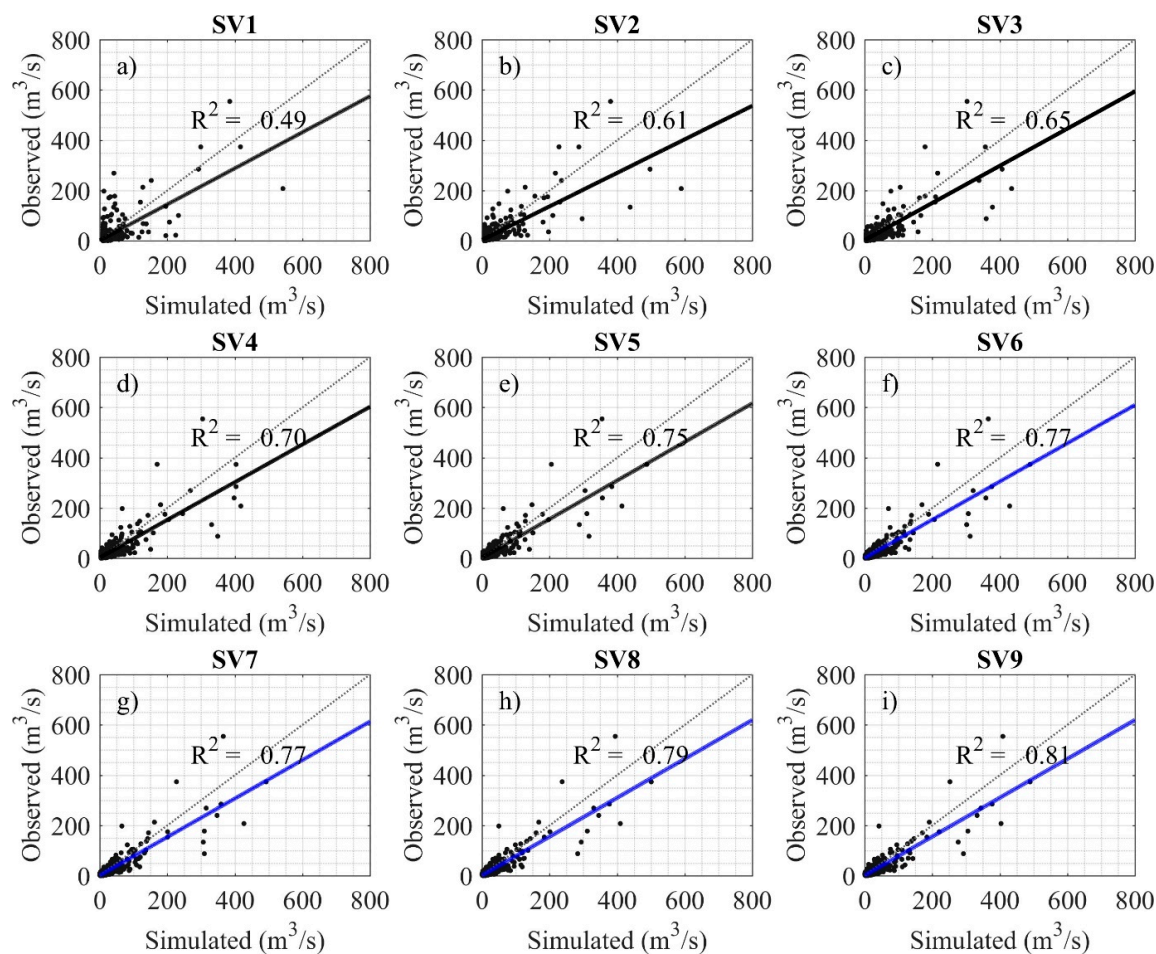


Figure 10. SV1–SV9 model scatter plots with the corresponding linear regression lines and R^2 over the testing period.

Table 4. Tank-LSSVM model performance measures using different input combinations.

Model	Input Combinations	Training (2007–2013)				Testing (2014–2016)			
		NSE	R^2	RMSE	RMSE Q70	NSE	R^2	RMSE	RMSE Q70
Tank		0.92	0.96	20.18	3.74	0.81	0.91	14.72	3.12
HY1	$P_{(t)}, ST1_{(t)}$	0.92	0.96	19.74	4.19	0.75	0.80	16.63	3.11
HY2	$P_{(t)}, ST1_{(t)}, ST2_{(t)}$	0.93	0.96	18.49	2.99	0.76	0.80	16.52	2.67
HY3	$P_{(t)}, ST1_{(t)}, ST2_{(t)}, ST3_{(t)}$	0.95	0.97	16.24	2.50	0.85	0.86	12.91	2.13
HY4	$P_{(t)}, ST1_{(t)}, ST2_{(t)}, ST3_{(t)}, \theta_{(t)}$	0.94	0.97	17.43	2.34	0.85	0.85	12.96	2.23
HY5	$P_{(t)}, P_{(t-1)}, ST1_{(t)}$	0.92	0.96	19.87	4.40	0.71	0.77	17.93	3.11
HY6	$P_{(t)}, ST1_{(t)}, ST1_{(t-1)}, ST2_{(t)}, ST3_{(t)}$	0.93	0.96	18.75	2.94	0.84	0.85	13.38	2.19
HY7	$P_{(t)}, ST1_{(t)}, ST2_{(t)}, ST2_{(t-1)}, ST3_{(t)}$	0.93	0.96	18.87	2.99	0.85	0.85	13.18	2.34
HY8	$P_{(t)}, ST1_{(t)}, ST2_{(t)}, ST3_{(t)}, ST3_{(t-1)}$	0.93	0.96	18.97	2.93	0.85	0.85	13.13	2.39
HY9	$P_{(t)}, ST1_{(t)}, ST2_{(t)}, ST3_{(t)}, \theta_{(t)}, \theta_{(t-1)}$	0.93	0.96	18.98	2.81	0.85	0.85	13.14	2.49

The Tank-LSSVM models' NSE were all classified as “very good” based on their performance during the training and testing periods. Compared to the individual LSSVM and tank models, improved daily runoff simulations were seen in the Tank-LSSVM model. The HY2 model's overall performance showed a considerable improvement over the HY1 model in terms of the RMSE Q70 low flow simulation. This was evident during both the training and testing periods, where the HY2 model results showed a considerable reduction in RMSE (i.e., 2.67 m³/s) with respect to the tank

model's $3.12 \text{ m}^3/\text{s}$. We concluded that the intermediate SM states inferred from the 2nd tank storage ST2 played an essential role in simulating low flow in the HY2 model. Including the 3rd tank storage in the HY3 model substantially improved the low flow simulation, where the RMSE Q70 was $2.13 \text{ m}^3/\text{s}$. These results were comparable to the performance metrics in the models HY1 and HY2. Similar to the 2nd tank storage's role in HY2, the 3rd tank storage played a critical role in describing HY3's base flow as a proxy variable. This led to a substantial improvement in the low flow simulation. The Tank-LSSVM models' performances were also confirmed by graphical representation, as displayed in Figure 11. A linear regression of the HY4 model showed that the observed runoff was underestimated by about 6% (i.e., $y = 0.94x + 0.5$), which appeared to fail to provide a visible improvement in the RR simulation when adding satellite SM products into the proposed LSSVM-based RR framework. In other words, the satellite SM product contribution appeared insignificant. This was partially due to the fact that the SM temporal dynamics were described by the tank storage.

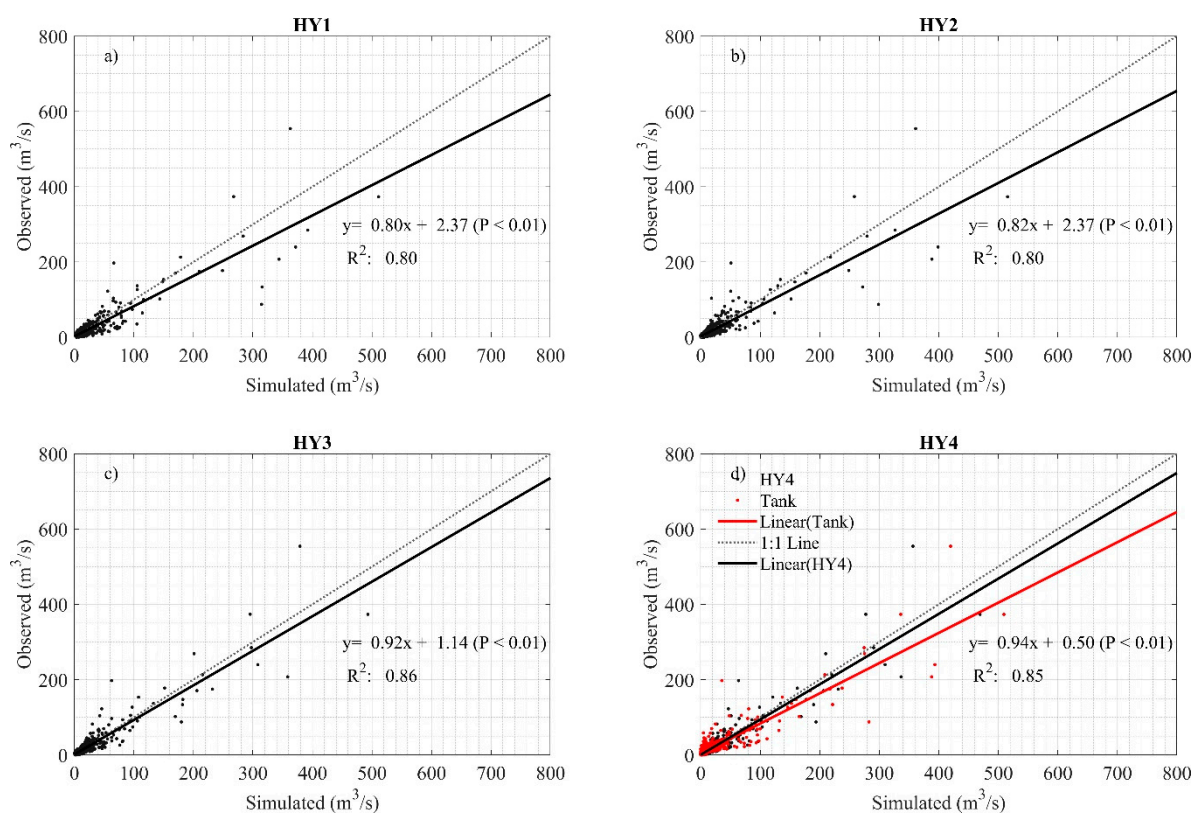


Figure 11. Scatterplots of the observed and simulated runoff over the testing period. The dotted lines represent perfect linear relationships, while the solid black lines indicate a linear fit between the simulated and observed runoff. The result of the tank model simulation is displayed with the solid red line in panel (d).

Accurate RR modeling is of great importance during dry seasons for water resource planning and management studies, especially for water quality and drought issues associated with low flows. To assess the accuracy of low flow simulations, flow duration curves derived from the simulated runoff were compared to the observed runoff from 2014 to 2016. Apart from the HY1 model, the Tank-LSSVM model's low flow simulations were better than those of the tank model (Figure 12). Figure 13 shows the simulated runoff time series from the tank model and the Tank-LSSVM model, along with the observed runoff data. The figure indicates that rapid low flow fluctuations are better simulated by the Tank-LSSVM model.

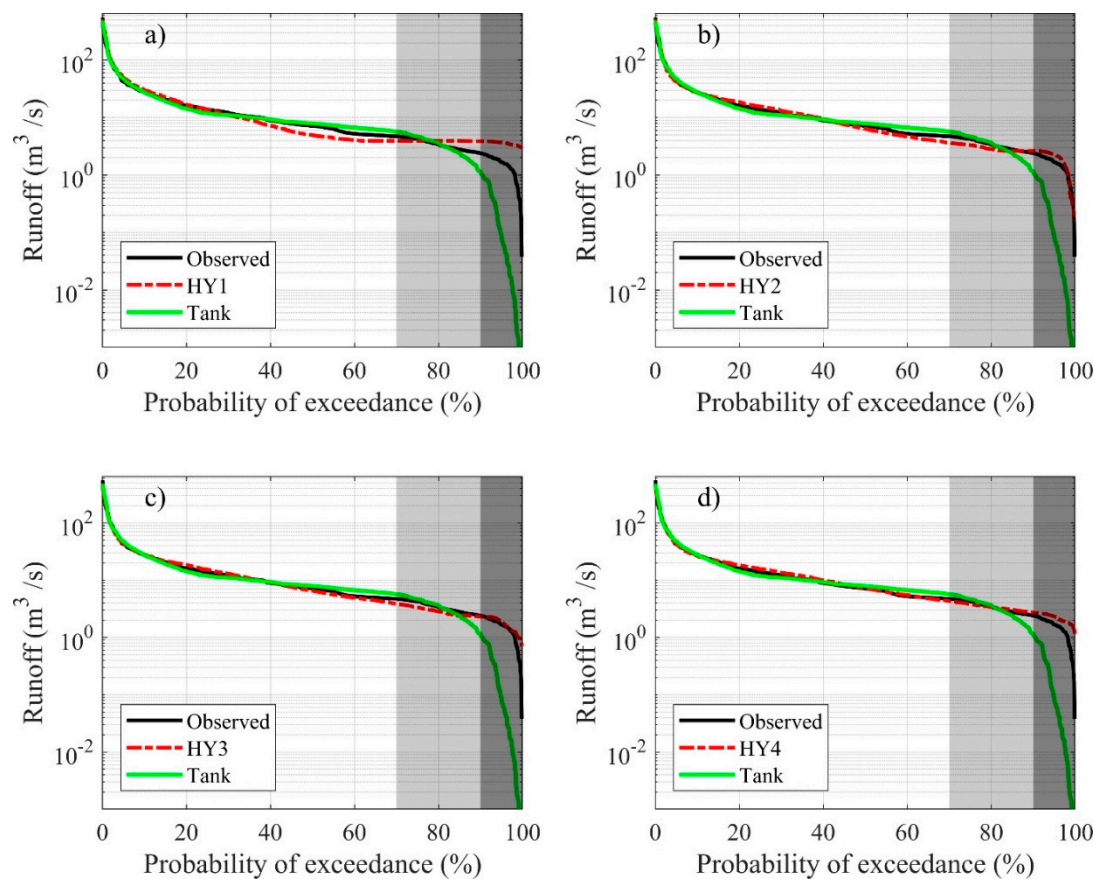


Figure 12. Flow duration curve comparison between four different Tank-LSSVM models, the tank model, and direct observation data during the testing period (2014–2016). Here, the light and dark shaded areas represent 70% and 90% exceedance percentiles of the observed runoff.

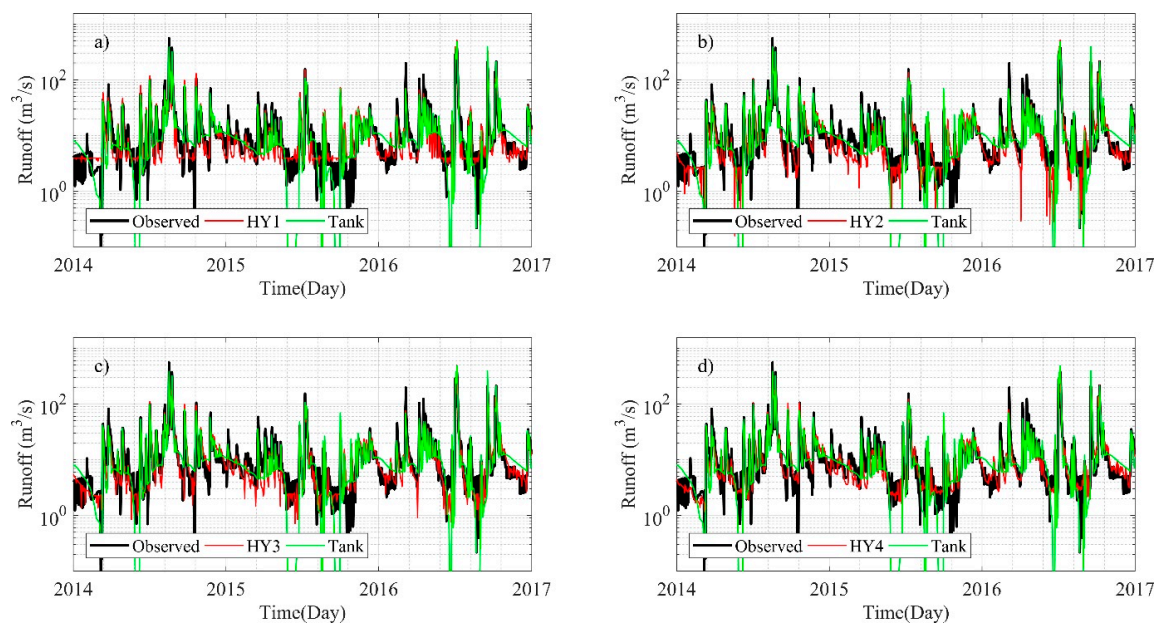


Figure 13. Simulated runoff comparison between the tank model and four different Tank-LSSVM models (HY1, HY2, HY3, and HY4). Direct observation runoff data from 2014 to 2016 are also included.

5. Concluding Remarks

In this study, we explored a new RR model that combined intermediate state variables obtained from a tank model with an LSSVM-based nonlinear regression model. The main study assumption was that combining the different models would be more favorable and efficient for runoff modeling than using each model individually. The hybrid RR model's performance was compared to two individual RR models—the tank model and the LSSVM model. The main findings are summarized as follows:

(1) The tank model's performance with the calibrated parameters confirms that it is capable of accurately describing rainfall-runoff relationships and is categorized as “very good ($NSE > 0.7$)”. The tank model shows relatively large deviations, however, for low flow simulations, suggesting that the tank model alone is insufficient for simulating particular RR processes.

(2) This study first explored the LSSVM-based RR model without the intermediate variables from the tank model. The LSSVM-based RR model with rainfall and ESA CCI_{SWI} lagged predictor input values demonstrated that the satellite SM data were effective for describing the SM's temporal dynamics. The LSSVM model's performance is classified as “good ($0.5 \leq NSE < 0.7$)” or “very good ($NSE > 0.7$)”, depending on different combinations of time-lagged input variables. Although the overall performance of the LSSVM model alone is generally lower than that of the tank model, the results support satellite-based SM product use for hydrological applications.

(3) The Tank-LSSVM models' NSE are all classified as “very good” based on their performance during the training and testing periods. Compared to the individual LSSVM and tank models, improved daily runoff simulations are seen in the Tank-LSSVM model. In particular, the Tank-LSSVM model including the intermediate state variables has considerably improved low flow simulation during the training and testing periods. The improvement of the LSSVM over the tank model may be partially due to the time-lagged input vectors that represent the routing effect in the rainfall-runoff process. The satellite SM products have not substantially contributed to the low flow simulations because the SM's temporal dynamics are largely described by tank storage. The results confirm that the SM state variables derived from the well-calibrated continuous RR model can better represent the SM's temporal dynamics than those obtained from the satellite SM data.

Due to the runoff simulation improvements in the hybrid RR model, this study's modeling framework could be beneficial and relevant for a number of different hydrologic applications. To support this study's findings, future work is necessary to explore other conceptual models and physically-based models for different regions with longer records. Combining existing RR models with an LSSVM-based RR framework outperforms models that use satellite SM data. This is partially because the existing RR models use the mean value over the catchment area rather than spatially distributed SM. This can lead to a misrepresentation of overall performance in the proposed modeling the framework. Moreover, we acknowledge that the obtained results from the tank model could be affected by the optimization scheme used in the calibration process. More specifically, the gradient-based automatic optimization scheme could be possibly trapped in local optima that are far from the desired global optima. Future work will further investigate the model sensitivity with different optimization schemes.

Author Contributions: We jointly conceived the project and designed the study. M.K. and H.-H.K. wrote the majority of the manuscript. H.-H.K. and D.H. critically revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Korea Meteorological Administration Research and Development Program under Grant KMI (KMI2018-07010).

Acknowledgments: Climate data (i.e., air temperature, relative humidity, wind speed, and hours of sunshine) are freely available from <https://data.kma.go.kr>. Runoff and rainfall data can be obtained from the K-water and can be downloaded from <http://www.wamis.go.kr/>. The ESA CCI SM products (v04.2) used in this study are obtained from <https://www.esa-soilmoisture-cci.org/>.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Tank models with vertically interconnected tanks (three tanks for this study) simulate RR processes, such as flood events and continuous runoff. In the 3-tank model, the side outlets in the 1st tank represent the surface runoff (q_{12} and q_{11}), while the outlets in the 2nd and 3rd tank are considered the intermediate runoff (q_2) and base flow (q_3), respectively. The outputs of each side outlet are calculated based on the following formulas:

1st tank:

$$\frac{dST_1}{dt} = P - ET - q_{12} - q_{11} - I_1 \quad (A1)$$

$$\text{if } ST_1 > H_{12}, q_{12} = (ST_1 - H_{12}) \times a_{12}; \text{ if } ST_1 \leq H_{12}, q_{12} = 0 \quad (A2)$$

$$\text{if } ST_1 > H_{11}, q_{11} = (ST_1 - H_{11}) \times a_{11}; \text{ if } ST_1 \leq H_{11}, q_{11} = 0 \quad (A3)$$

$$I_1 = ST_1 \times b_1 \quad (A4)$$

2nd tank:

$$\frac{dST_2}{dt} = I_1 - I_2 - q_2 \quad (A5)$$

$$\text{if } ST_2 > H_2, q_2 = (ST_2 - H_2) \times a_2; \text{ if } ST_2 \leq H_2, q_2 = 0 \quad (A6)$$

$$I_2 = ST_2 \times b_2 \quad (A7)$$

3rd tank:

$$\frac{dST_3}{dt} = I_2 - q_3 \quad (A8)$$

$$q_3 = ST_3 \times a_3 \quad (A9)$$

Total runoff:

$$Q = q_{11} + q_{12} + q_2 + q_3 \quad (A10)$$

where $\frac{dS}{dt}$ is the time-dependent tank storage. P , ET , and I refer to rainfall, evapotranspiration, and infiltration, respectively.

References

- Curran, J.H. Streamflow Record Extension for Selected Strams in the Susitan River Basin, Alaska. *US Geol. Surv. Sci. Investig. Rep* **2012**, *5210*, 36.
- Sittner, W.T. WMO project on intercomparison of conceptual models used in hydrological forecasting. *Hydrol. Sci. Bull.* **1976**, *21*, 203–213. [\[CrossRef\]](#)
- Sivapragasam, C.; Liong, S.; Pasha, M. Rainfall and runoff forecasting with SSA-SVM approach. *J. Hydroinform.* **2001**, *3*, 141–152. [\[CrossRef\]](#)
- Zhuo, L.; Han, D. Could operational hydrological models be made compatible with satellite soil moisture observations? *Hydrol. Process.* **2016**, *30*, 1637–1648. [\[CrossRef\]](#)
- McIntyre, N.; Lee, H.; Wheeler, H.; Young, A.; Wagener, T. Ensemble predictions of runoff in ungauged catchments. *Water Resour. Res.* **2005**, *41*, 1–14. [\[CrossRef\]](#)
- Jakeman, A.J. How Much Complexity Is Warranted in a Rainfall-Runoff Model? are good predictors of streamflow and. *Water Resour. Res.* **1993**, *29*, 2637–2649. [\[CrossRef\]](#)
- Pfannerstill, M.; Guse, B.; Fohrer, N. Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. *J. Hydrol.* **2014**, *510*, 447–458. [\[CrossRef\]](#)
- Orth, R.; Staudinger, M.; Seneviratne, S.I.; Seibert, J.; Zappa, M. Does model performance improve with complexity? A case study with three hydrological models. *J. Hydrol.* **2015**, *523*, 147–159. [\[CrossRef\]](#)
- Kisi, O.; Parmar, K.S. Application of multivariate adaptive regression spline models in long term prediction of river water pollution. *J. Hydrol.* **2016**, *534*, 104–112. [\[CrossRef\]](#)
- Devia, G.K.; Ganasri, B.P.; Dwarakish, G.S. A Review on Hydrological Models. *Aquat. Proced.* **2015**, *4*, 1001–1007. [\[CrossRef\]](#)

11. Behzad, M.; Asghari, K.; Eazi, M.; Palhang, M. Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Syst. Appl.* **2009**, *36*, 7624–7629. [\[CrossRef\]](#)
12. Young, C.C.; Liu, W.C.; Wu, M.C. A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. *Appl. Soft Comput. J.* **2017**, *53*, 205–216. [\[CrossRef\]](#)
13. Song, J.H.; Her, Y.; Park, J.; Lee, K.D.; Kang, M.S. Simulink Implementation of a Hydrologic Model: A Tank Model Case Study. *Water* **2017**, *9*, 639. [\[CrossRef\]](#)
14. Paik, K.; Kim, J.H.; Kim, H.S.; Lee, D.R. A conceptual rainfall-runoff model considering seasonal variation. *Hydrol. Process.* **2005**, *19*, 475–476. [\[CrossRef\]](#)
15. Beven, K.J.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [\[CrossRef\]](#)
16. Burnash, R.J.; Ferral, R.L.; McGuire, R.A. *A Generalized Streamflow Simulation System: Conceptual Models for Digital Computers*; Joint Federal State River Forecast Center: Sacramento, CA, USA, 1973.
17. Moore, R.J. The probability-distributed principle and runoff production at point and basin scales. *Hydrol. Sci. J.* **1985**, *30*, 273–297. [\[CrossRef\]](#)
18. Bergström, S. Development and Application of a Conceptual Runoff Model for Scandinavian Catchments. *Smhi* **1976**, *RHO 7*, 134. [\[CrossRef\]](#)
19. Graham, D.N.; Butts, M.B. Flexible Integrated Watershed Modeling with MIKE SHE. In *Watershed Models*; Singh, V.P., Donald, K.F., Eds.; CRC Press: Boca Raton, FL, USA, 2005; pp. 245–272. ISBN 0849336090.
20. Neitsch, S.; Arnold, J.; Kiniry, J.; Williams, J. *Soil and Water Assessment Tool Theoretical Documentation Version 2009*; Texas Water Resources Institute: College Station, TX, USA, 2011.
21. Vaze, J.; Jordan, P.; Beecham, R.; Frost, A.; Summerell, G. *Guidelines for Rainfall-Runoff Modelling: Towards Best Practice Model Application*; eWater Cooperative Research Centre: Australia, 2011; ISBN 9781921543517.
22. Sugawara, M. Automatic calibration of the tank model. *Hydrol. Sci. Bull.* **1979**, *24*, 375–388. [\[CrossRef\]](#)
23. Basri, H. Development of Rainfall-runoff Model Using Tank Model: Problems and Challenges in Province of Aceh, Indonesia. *Aceh Int. J. Sci. Technol.* **2013**, *2*, 26–36. [\[CrossRef\]](#)
24. Fumikazu, N.; Toshisuke, M.; Yoshio, H.; Hiroshi, T.; Kimihito, N. Evaluation of water resources by snow storage using water balance and tank model method in the Tadori River basin of Japan. *Paddy Water Environ.* **2013**, *11*, 113–121. [\[CrossRef\]](#)
25. Samui, P.; Kothari, D.P. Utilization of a least square support vector machine (LSSVM) for slope stability analysis. *Sci. Iran.* **2011**, *18*, 53–58. [\[CrossRef\]](#)
26. Bray, M.; Han, D. Identification of support vector machines for runoff modelling. *J. Hydroinform.* **2004**, 265–280. [\[CrossRef\]](#)
27. Granata, F.; Gargano, R.; de Marinis, G. Support Vector Regression for Rainfall-Runoff Modeling in Urban Drainage: A Comparison with the EPA's Storm Water Management Model. *Water* **2016**, *8*, 69. [\[CrossRef\]](#)
28. Wu, M.C.; Lin, G.F.; Lin, H.Y. Improving the forecasts of extreme streamflow by support vector regression with the data extracted by self-organizing map. *Hydrol. Process.* **2014**, *28*, 386–397. [\[CrossRef\]](#)
29. Fernando, A.K.; Shamseldin, A.Y.; Abrahart, R.J. Use of Gene Expression Programming for Multimodel Combination of Rainfall-Runoff Models. *J. Hydrol. Eng.* **2012**, *17*, 975–985. [\[CrossRef\]](#)
30. Hosseini, S.M.; Mahjouri, N. Integrating Support Vector Regression and a geomorphologic Artificial Neural Network for daily rainfall-runoff modeling. *Appl. Soft Comput. J.* **2016**, *38*, 329–345. [\[CrossRef\]](#)
31. Okkan, U.; Serbes, Z.A. Rainfall-runoff modeling using least squares support vector machines. *Environmetrics* **2012**, *23*, 549–564. [\[CrossRef\]](#)
32. Massari, C.; Brocca, L.; Ciabatta, L.; Moramarco, T.; Gabellani, S.; Albergel, C.; De Rosnay, P.; Puca, S.; Wagner, W. The Use of H-SAF Soil Moisture Products for Operational Hydrology: Flood Modelling over Italy. *Hydrology* **2015**, *2*, 2–22. [\[CrossRef\]](#)
33. Dorigo, W.; Wagner, W.; Albergel, C.; Albrecht, F.; Balsamo, G.; Brocca, L.; Chung, D.; Ertl, M.; Forkel, M.; Gruber, A.; et al. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* **2017**, *203*, 185–215. [\[CrossRef\]](#)
34. Loizu, J.; Massari, C.; Álvarez-Mozos, J.; Tarpanelli, A.; Brocca, L.; Casali, J. On the assimilation set-up of ASCAT soil moisture data for improving streamflow catchment simulation. *Adv. Water Resour.* **2018**, *111*, 86–104. [\[CrossRef\]](#)

35. Brocca, L.; Melone, F.; Moramarco, T.; Wager, W.; Naeimi, V.; Bartalis, Z.; Hasenauer, S. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 1881–1893. [\[CrossRef\]](#)
36. Dharssi, I.; Bovis, K.J.; Macpherson, B.; Jones, C.P. Operational assimilation of ASCAT surface soil wetness at the Met Office. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 2729–2746. [\[CrossRef\]](#)
37. Ciabatta, L.; Massari, C.; Brocca, L.; Gruber, A.; Reimer, C.; Hahn, S.; Paulik, C.; Dorigo, W.; Kidd, R.; Wagner, W. SM2RAIN-CCI: A new global long-term rainfall data set derived from ESA CCI soil moisture. *Earth Syst. Sci. Data* **2018**, *10*, 267–280. [\[CrossRef\]](#)
38. Brocca, L.; Ciabatta, L.; Massari, C.; Moramarco, T.; Hahn, S.; Hasenauer, S.; Kidd, R.; Dorigo, W.; Wagner, W.; Levizzani, V. Journal of Geophysical Research: Atmospheres rainfall from satellite soil moisture data. *J. Geophys. Res.* **2014**, 1–14. [\[CrossRef\]](#)
39. Enenkel, M.; Steiner, C.; Mistelbauer, T.; Dorigo, W.; Wagner, W.; See, L.; Atzberger, C.; Schneider, S.; Rogenhofer, E. A combined satellite-derived drought indicator to support humanitarian aid organizations. *Remote Sens.* **2016**, *8*, 340. [\[CrossRef\]](#)
40. Brocca, L.; Moramarco, T.; Melone, F.; Wagner, W.; Hasenauer, S.; Hahn, S. Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall-runoff modeling. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2542–2555. [\[CrossRef\]](#)
41. Lievens, H.; Tomer, S.K.; Al Bitar, A.; De Lannoy, G.J.M.; Drusch, M.; Dumedah, G.; Hendricks Franssen, H.J.; Kerr, Y.H.; Martens, B.; Pan, M.; et al. SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sens. Environ.* **2015**, *168*, 146–162. [\[CrossRef\]](#)
42. Massari, C.; Brocca, L.; Tarpanelli, A.; Moramarco, T. Data assimilation of satellite soil moisture into rainfall-runoff modelling: A complex recipe? *Remote Sens.* **2015**, *7*, 11403–11433. [\[CrossRef\]](#)
43. Han, E.; Merwade, V.; Heathman, G.C. Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. *J. Hydrol.* **2012**, *416–417*, 98–117. [\[CrossRef\]](#)
44. Yoo, J.H. Maximization of hydropower generation through the application of a linear programming model. *J. Hydrol.* **2009**, *376*, 182–187. [\[CrossRef\]](#)
45. Topp, G.C.; Davis, J.L.; Annan, A.P. Electromagnetic Determination of Soil Water Content: Measurements in Coaxial Transmission Lines. *Water Resour. Res.* **1980**, *16*, 574–582. [\[CrossRef\]](#)
46. Dorigo, W.A.; Gruber, A.; De Jeu, R.A.M.; Wagner, W.; Stacke, T.; Loew, A.; Albergel, C.; Brocca, L.; Chung, D.; Parinussa, R.M.; et al. Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sens. Environ.* **2015**, *162*, 380–395. [\[CrossRef\]](#)
47. Liu, Y.Y.; Dorigo, W.A.; Parinussa, R.M.; De Jeu, R.A.M.; Wagner, W.; McCabe, M.F.; Evans, J.P.; Van Dijk, A.I.J.M. Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sens. Environ.* **2012**, *123*, 280–297. [\[CrossRef\]](#)
48. Liu, Y.Y.; Parinussa, R.M.; Dorigo, W.A.; De Jeu, R.A.M.; Wagner, W.; Van Dijk, M.A.I.J.; McCabe, M.F.; Evans, J.P. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 425–436. [\[CrossRef\]](#)
49. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M.; Ab, W. *Crop Evapotranspiration—Guidelines for Computing Reference Crop Evapotranspiration*; FAO: Roma, Italy, 1998; pp. 1–15.
50. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
51. Raghavendra, S.; Deka, P.C. Support vector machine applications in the field of hydrology: A review. *Appl. Soft Comput. J.* **2014**, *19*, 372–386. [\[CrossRef\]](#)
52. Yan, X.; Chowdhury, N.A. Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach. *Int. J. Electr. Power Energy Syst.* **2013**, *53*, 20–26. [\[CrossRef\]](#)
53. Suykens, J.A.K.; De Brabanter, J.; Lukas, L.; Vandewalle, J. Weighted least squares support vector machines: Robustness and sparse approximation. *Neurocomputing* **2002**, *48*, 85–105. [\[CrossRef\]](#)
54. Xavier-De-Souza, S.; Suykens, J.A.K.; Vandewalle, J.; Bolle, D. Coupled simulated annealing. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2010**, *40*, 320–335. [\[CrossRef\]](#)
55. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [\[CrossRef\]](#)
56. Yu, P.S.; Chen, S.T.; Chang, I.F. Support vector regression for real-time flood stage forecasting. *J. Hydrol.* **2006**, *328*, 704–716. [\[CrossRef\]](#)

57. Massari, C.; Brocca, L.; Barbetta, S.; Papathanasiou, C.; Mimikou, M.; Moramarco, T. Using globally available soil moisture indicators for flood modelling in Mediterranean catchments. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 839–853. [[CrossRef](#)]
58. Massari, C.; Camici, S.; Ciabatta, L.; Brocca, L. Exploiting satellite-based surface soil moisture for flood forecasting in the Mediterranean area: State update versus rainfall correction. *Remote Sens.* **2018**, *10*, 292. [[CrossRef](#)]
59. Silvestro, F.; Gabellani, S.; Rudari, R.; Delogu, F.; Laiolo, P.; Boni, G. Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 1727–1751. [[CrossRef](#)]
60. Albergel, C.; Rüdiger, C.; Pellarin, T.; Calvet, J.-C.; Fritz, N.; Froissard, F.; Suquia, D.; Petitpa, A.; Pignatelli, B.; Martin, E. From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. *Hydrol. Earth Syst. Sci. Discuss.* **2008**, *5*, 1603–1640. [[CrossRef](#)]
61. Wagner, W.; Lemoine, G.; Rott, H. A method for estimating soil moisture from ERS Scatterometer and soil data. *Remote Sens. Environ.* **1999**, *70*, 191–207. [[CrossRef](#)]
62. Legates, D.R.; McCabe, G.J. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241. [[CrossRef](#)]
63. Kalin, L.; Isik, S.; Schoonover, J.E.; Lockaby, B.G. Predicting Water Quality in Unmonitored Watersheds Using Artificial Neural Networks. *J. Environ. Qual.* **2010**, *39*, 1429. [[CrossRef](#)]
64. Bober, W. *Introduction to Numerical and Analytical Methods with MATLAB for Engineers and Scientists*; CRC Press: Boca Raton, FL, USA, 2013; ISBN 9781466576094.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).